

# SYNTHESIS OF HYBRID POLYNUCLEOTIDE MOLECULES USING SINGLE-STRANDED POLYNUCLEOTIDE MOLECULES

## BACKGROUND OF THE INVENTION

### 1. Field of the Invention

The present invention relates to a method for the production of hybrid polynucleotide molecules from two parental polynucleotide molecules using single-stranded polynucleotides, which can achieve superior effects to the existing methods, such as a higher frequencies.

### 2. Description of the Related Art

DNA shuffling is a developed technique that allows accelerated and directed protein evolution *in vitro*. In this method, the acquisition of genes encoding improved proteins is done in two steps. In the first step, a single gene is mutagenized, and desired mutant genes are selected. In the second step, the mutant genes are fragmented by DNase I, and subsequently recombined *in vitro* by using PCR. Among the recombinants, (i.e., the products of DNA shuffling), those producing most favored proteins are isolated (Fig. 1; 1, 2). The modified versions of the DNA shuffling exist: (i) random priming was used to generate DNA fragment instead of the DNase I digestion (3); (ii) PCR conditions with very short annealing/extension steps were employed to increase the frequency of

recombination (4). Using these techniques, a number of improved enzymes have been obtained (1-9). When the DNA shuffling is done using a set of homologous genes instead of a set of mutant genes derived from a single gene, this technique is called family shuffling. Family shuffling utilizes naturally occurring nucleotide substitutions among family genes as the driving force for the *in vitro* evolution. The application of the family shuffling strategy has also provided many successful examples (10-15).

A potential problem of the family shuffling is a low yield of recombinants (i.e., hybrid molecules constituted from several family gene sequences). When two parental genes of 80% nucleotide sequence identity were shuffled, the frequency of hybrid formation was less than 1% (16-17). The low recombination yield may be due to a lower frequency of the heteroduplex formation compared to the frequency of the homoduplex formation (Fig. 2; 16).

Thus, it is desirable to develop methods which allow for the high frequency production of recombinant molecules of two or more family genes. We have reported a technique which is the shuffling of restriction endonuclease-digested DNA fragments instead of the shuffling of randomly fragmented DNA (16). The annealing of endonuclease-digested DNA fragments would produce homoduplex at a high

frequency, but significant DNA elongation only occurs on the heteroduplex molecules.

#### SUMMARY OF THE INVENTION

Contemporary genes belonging to the same gene family are derived from a single ancestral gene after repeated introduction of mutations through the natural divergent evolution processes (Fig. 3). The shuffling of the family gene sequences creates a library of chimeric genes which would yield gene products of diverse properties. Among the diverse properties expressed from the chimeric genes, desired ones could be selected.

In the shuffling of family genes in conventional methods (1-11, 13-15, 18), two types of annealing occur: homoduplex (annealing of strands derived from the same gene) and heteroduplex (annealing of strands derived from two different genes) (Fig. 2). If the frequency of the homoduplex formation is higher than that of the heteroduplex formation, the frequency of the regeneration of the original gene sequences through PCR reactions occurs preferentially rather than the formation of chimeric genes. The consequence of it is a low yield of recombinant molecules of two or more family genes.

Now, we invented a new technique which uses single-stranded polynucleotide sequences (17). In this method, single-stranded polynucleotides of two family genes are

prepared. One single-stranded polynucleotide is the coding strand of one gene while another single-stranded polynucleotides is the non-coding strand of another gene. These two single-stranded polynucleotide molecules are fragmented into appropriate sizes, and their fragments are used for the family shuffling. This technique is preferable in that, in the first round of hybridization, the homoduplex formation is prevented.

Based on this new technique, we completed the present invention which can provide hybrid polynucleotide molecules from parental polynucleotides containing one or more parts of homology and one or more parts of heterology (e.g., family genes).

In this invention, two types of single-stranded polynucleotides are prepared from two parent polynucleotids (for example, two family genes, two genes obtained by random mutagenesis of a parental gene, etc.), and used for the *in vitro* polynucleotide synthesis. The first type is the single-stranded polynucleotide molecule corresponding to the coding strand of, for example, one family gene while the second type is the single-stranded molecule corresponding to the non-coding strand of, for example, another family gene. In other words, these two types of the single-stranded polynucleotide molecules are complementary to each other but have one or more parts

which are not complementary (the first-type molecule comprises one or more parts of homology and one or more parts of heterology to the complementary sequence of the second-type molecule). These single-stranded polynucleotide molecules are fragmented into appropriate sizes, and hybridized. In this procedure, no homoduplex molecule is formed, and when appropriate enzyme(s) exists, the nucleotide elongation can occur on heteroduplex molecules forming chimeric polynucleotide pieces. (Fig. 4) Subsequently, the steps of the denaturation followed by hybridization and polynucleotide synthesis could be done repeatedly as summarized in Fig. 1.

This method can be combined to appropriate methods for introducing one or more mutations into chimeric genes.

Thus, the present invention provides *inter alia*:

(1) A method for making libraries of hybrid polynucleotide molecules in which double-stranded polynucleotide molecules are not used as starting materials.

(2) The method of (1) above, wherein two types of single-stranded polynucleotide molecules are used as starting materials and wherein the first-type molecule comprises stretches of sequences containing one or more parts of homology and one or more parts of heterology to the complementary sequence of the second-type molecule.

(3) The method of (2) above, wherein the single-stranded polynucleotide molecules are fragmented and used as templates for *de novo* polynucleotide synthesis to create hybrid polynucleotide molecules.

(4) The method of (2), wherein mutations are introduced into hybrid polynucleotide molecules prior, during or after the production of the hybrid polynucleotide molecules.

(5) A method for making libraries of hybrid polynucleotide molecules, which comprises:

(i) preparing two single-stranded polynucleotide molecules comprising sequences which are complementary to each other,

(ii) randomly or non-randomly fragmenting the two single-stranded polynucleotide molecules,

(iii) incubating the fragmented molecules under conditions such that hybridization of fragmented polynucleotide molecules occurs and *de novo* polynucleotide synthesis on the hybridized molecules occurs,

(iv) denaturing the resultant elongated double-stranded polynucleotide molecules into single-stranded polynucleotide molecules,

(v) incubating the resultant single-stranded polynucleotide molecules under conditions such that hybridization of single-stranded polynucleotide molecules

occurs and de novo polynucleotide synthesis on the hybridized molecules occurs, and

(vi) repeating at least two further cycles of steps (iv) and (v).

#### BRIEF DESCRIPTION OF THE DRAWINGS

A more complete appreciation of the invention and many of the attendant advantages thereof will be readily obtained as the same become better understood by reference to the following detailed description when considered in connection with the accompanying drawings, wherein:

Fig. 1 illustrates a concept of DNA shuffling. The method for *in vitro* evolution of genes generally called DNA shuffling consists of repeated cycles of two steps: random mutagenesis and DNA shuffling of mutations. In the first step (random mutagenesis/screening), several mutants are isolated from a target gene. In the second step, the mutant genes are segmented (usually by DNase I), then reassembled by PCR without primer. The reassembled molecules are further amplified by PCR with appropriate primers. DNAs created by DNA shuffling are used to transform a recipient (usually *Escherichia coli*), and transformed strains are selected for desired phenotypes. Instead of the DNase I segmentation, random-priming synthesis of short fragments can be used.

Fig. 2 illustrates formation of homoduplex and heteroduplex from two parental polynucleotide molecules whose sequences are similar. When two parental polynucleotide molecules whose sequences are similar but not identical to each other are denatured and hybridized, two types of hybridized molecules are formed: one strand of polynucleotide molecule hybridized with a complementary strand derived from the same parent (homoduplex), and that hybridized with a complementary strand derived from a different parent (heteroduplex). The velocity constant for the homoduplex formation ( $k_1$ ) may be higher than that for the heteroduplex formation ( $k_2$ ).

Fig. 3 illustrates a concept of family shuffling. Contemporary genes belonging to the same gene family are derived from a single ancestral gene after repeated introduction of mutations through the natural divergent evolution processes. The shuffling of the family gene sequences creates a library of chimeric genes from which desired chimera are selected.

Fig. 4A and 4B illustrates hybridization of polynucleotides and nucleotide synthesis in family shuffling.

4A: Two polynucleotide strands (e.g. DNA) of two family genes are indicated by open and close boxes. In the family shuffling using two genes, two types of annealing



occur: homoduplex (annealing of strands derived from the same gene) and heteroduplex (annealing of strands derived from two different genes). If the frequency of the homoduplex formation is higher than that of the heteroduplex formation, the frequency of the regeneration of the original (parental) gene sequences occurs preferentially rather than the formation of chimeric genes.

4B: To prevent the homoduplex formation in the first round of nucleotide synthesis, single-stranded polynucleotide molecules are prepared from two family genes. These single-stranded polynucleotide molecules are fragmented by appropriate methods (e.g. by DNase I digestion), and hybridized. In contrast to conventional methods using double-stranded polynucleotide molecules, only heteroduplex molecules are formed in this invention.

Fig. 5 illustrates chimeric structures of the single-stranded DNA-based shuffling products. This is an example where single-stranded DNAs of *nahH* and *xyle* were shuffled, fifty shuffled clones exhibiting the C230 activity were randomly selected, and their nucleotide sequences were determined as described in Example below. The sequences derived from *nahH* and *xyle* are shown by shaded and solid boxes, respectively. White boxes show the expected recombination regions, and solid triangles show the point mutations.

### DETAILED DESCRIPTION OF THE INVENTION

The term "parental polynucleotide molecule" is used to indicate a polynucleotide molecule species which is a starting material for *in vitro* manipulations.

The term "hybrid polynucleotide molecule" is used to indicate that the polynucleotide molecule is constituted from sequences derived from several parental polynucleotide molecules. The hybrid polynucleotide molecule may or may not contain base substitution(s) comparing to the parental sequences as mutation(s) can be introduced during the processes forming the hybrid polynucleotide molecule.

The term "initial template" is used to indicate that this molecule is used as a starting material for *in vitro* manipulations, and served as a template for the *de novo* polynucleotide synthesis.

The term "complementary" is used to mean that two polynucleotide sequences is homologous (not necessarily 100% identical) enough to make a stable hybridized molecules at a temperature above 0°C.

The term "random or non-random fragmentation" means the process to generate polynucleotide molecules shorter than the length of an initial template. The fragmentation can be attained by several means, for example, by either the digestion of a single-stranded polynucleotide by appropriate nuclease(s), a physical shearing (e.g.

sonication) of a polynucleotide molecule, or random priming of a polynucleotide to synthesize polynucleotide fragments complementary to the template sequence.

The term "*in vitro* manipulations" is used to mean various *in vitro* manipulations required for the synthesis of hybrid polynucleotide molecules starting from two or more types of homologous polynucleotide molecules.

The terms "homology" and "homologous" is used to indicate that the partial sequences of two or more polynucleotide molecules are either the same or similar to each other, or highly complementary to each other so that two complementary strands derived from two substantially homologous sequences can stably hybridize each other at a temperature above 0°C. The term "heterology" is used to indicate the opposite meaning.

The DNA shuffling methods (1-11, 13-15, 18) have been developed for accelerated and directed protein evolution *in vitro*. The methods generally consist of two steps: (i) the isolation of several improved mutants of the gene, and (ii) the *in vitro* recombination of the mutant genes followed by the selection of best progenies. In Fig. 1, these steps are illustrated. In the first step, a single gene is mutagenized by one of established methods, and desired mutant genes are selected. In the second step, the fragments of the mutant genes are obtained e.g. by DNase I

treatment or random priming, and the mixture of the fragments is subjected to PCR without primer. DNA pieces derived from different mutants are recombined in this PCR. Next, PCR was carried out with forward and reverse primers to amplify the full-length DNA segments.

Instead to use a set of mutant derivatives as shown in Fig. 1, a set of homologous genes could be shuffled to obtain genes synthesizing improved enzymes. This technique is called family shuffling. Family shuffling utilizes naturally occurring nucleotide substitutions among family genes as the driving force for the *in vitro* evolution. A potential problem of the family shuffling is a low yield of recombinants (i.e., hybrid molecules constituted from several family gene sequences). When two parental genes of 80% nucleotide sequence identity (*xyle* and *nahH*, see Examples) were shuffled, the frequency of hybrid formation was less than 1% (16-17).

Thus, it is desirable to develop methods which allow for the high frequency production of recombinant molecules of two or more family genes. The present invention relates to the DNA shuffling methods summarized in Fig. 1, but is different from them in which single-stranded polynucleotide molecules are used as starting materials instead of double-stranded polynucleotide molecules. As described below, this method allowed recombination of family genes, etc. at

a frequency higher than that in conventional DNA shuffling methods.

After the failure of the family shuffling in our hands, we considered possible reasons for it. For the successful generation of recombinant progeny of two family gene, A and B, the hybridization between single-stranded DNA of A and B (heteroduplex) should be formed, on which the elongation of DNA proceeds to form a recombined molecule of A and B (Fig. 4A). However, in this method, the hybridization between A and A or that between B and B also occur forming homoduplex molecules (Fig. 2). If the probability of the formation of heteroduplex molecules (a A strand hybridized with a B strand) is smaller than that of homoduplex molecules, i.e. a A (or B) strand hybridizing with another A (or B) strand (Fig. 2), the frequency of the hybrid molecules between A and B may be very low in the products of the family shuffling. A computer simulation demonstrated that this inference would be the case.

To obtain a high frequency recombination in the polynucleotide shuffling, we newly conceived to elevate the probability of the formation of heteroduplex molecules or reduce the probability of the formation of homoduplex molecules. In an attempt to increase the probability of the heteroduplex formation in the family shuffling of *xyle* and *nahH* (see below, Examples), the annealing temperature

in the PCR reactions was lowered. However, the majority of the shuffling products still had a structure of either *xyle* or *nahH* (i.e., the frequency of the formation of *xyle-nahH* hybrids was very low).

We interpreted these results as indicating that the probability of homologous pairing is much higher than that of heterologous pairing in the shuffling of two genes of modest homology (e.g. about 80% homology). We then attempted to decrease the probability of homoduplex formation. For the successful generation of recombinant progeny of two family genes, the hybridization between a single-stranded polynucleotide molecule of one family gene and a single stranded polynucleotide molecules of another family gene is required. However, in the conventional family shuffling methods, the hybridization of two strands derived from the same gene would occur preferentially over the hybridization of two strands each of which is derived from different family genes.

As a result of extensive studies, we newly found that use of single-stranded polynucleotide for shuffling can achieve the object of the present invention.

The present invention used single-stranded polymucleotides instead of double-stranded DNA as starting materials for preparing hybrid polynucleotide molecules from the parental polynucleotides. Thus, the present

invention provides a method for making libraries of hybrid polynucleotide molecules, a method for forming a mutagenized double-stranded polynucleotide, a method for obtaining a chimeric polynucleotide sequence, etc., wherein single-stranded polynucleotides instead of double-stranded DNA are used as starting materials.

One embodiment of the method of the present invention includes the following steps:

(i) preparation of two single-stranded polynucleotide molecules comprising parts of sequences which are complementary to each other,

(ii) random or non-random fragmentation of them,

(iii) hybridization of the fragmented molecules followed by polynucleotide synthesis (i.e. polynucleotide chain elongation) on the hybridized molecules,

(iv) separation of the chain elongation products (i.e. double-stranded polynucleotide molecules) into single-stranded polynucleotide molecules (denaturation),

(v) hybridization of the resultant single-stranded polynucleotide molecules followed by polynucleotide synthesis on the hybridized molecules, and

(vi) repeating at least two further cycles of step (iv) and (v) (e.g., under PCR conditions).

The present invention involves in the formation of hybrid polynucleotide molecules by mixing two parental

polynucleotide sequences. The partial homology should exist between two parental polynucleotide sequences so that the complementary strands derived from some regions of the two parental polynucleotide molecules can hybridize each other under conditions required for the *de novo* polynucleotide sequences. In other words, two complementary polynucleotide sequences have one or more parts of homology and one or more parts of heterology from each other. The required homology is dependent upon several parameters including the G-C contents of the parental molecules. The present invention is applicable even if the total homology is low, e.g., less than about 95%, and even less than about 80%. Preferably, each of the homologous regions (i.e., parts of homology) should be at least 15-base-long with the homology higher than 75%.

The total length of the two polynucleotide sequences are not particularly limited. Generally, about 30 bases to about 10,000 bases are preferable, and about 100 bases to about 2,000 bases are more preferable.

In the case where the mixing of more than two parental polynucleotide sequences is desired, hybrids of two parental sequences should be formed first, then, the single-stranded polynucleotide sequences of the hybrids should be prepared to mix with the third parental polynucleotide sequences.



Various methods existing for single-stranded DNA preparation can be utilized in the present invention. As an example of one method, the genes have to be cloned either in a phagemid vector or in a single-stranded DNA phage such as M13, and the single-stranded DNA have to be prepared from the filamentous phage particles. Two homologous genes contain regions separately cloned in phagemid vectors (for example, pBluescript) or single-stranded DNA phage vectors. The orientations of one gene with respect to the origin of the single-stranded DNA replication should be the same, while that of another gene should be opposite with respect to the origin of the single-stranded DNA replication. Then, the coding strand of one gene and the non-coding strand of another gene are synthesized and packaged in phage particles. In other words, single-stranded DNAs of the two homologous genes are complementary to each other.

The plasmids thus constructed are introduced into "male" *E. coli* such as JM109 (as the infection by single-stranded DNA phage requires F pili). If phagemids are used, cultures of the transformants should be infected with a helper phage such as VCS-M13 (Stratagene) or M13K07 (Pharmacia) to rescue of single-stranded DNA from the phagemids.

In supernatant of the culture, phagemid DNA packaged in phage capsules as well as helper phage are recovered. Single-stranded DNA is then prepared from the phage capsules reactions. The presence of the helper phage DNA does not interfere with the following reactions.

Other methods than using phagemid/phage vectors exist for the preparation of single-stranded DNA: asymmetric PCR is one of such techniques.

The single-stranded DNA obtained prepared above may contain a double-stranded DNA as a contaminant, but the amount of the double-stranded DNA is preferably controlled to 30% or less, more preferably 5% or less.

The single-stranded DNA is then fragmented by an appropriate method. The degree of fragmentation (i.e., length of fragmented polynucleotides) is not particularly limited. One skilled in the art can optionally decide the degree of fragmentation taking into consideration the full length of the polynucleotides to be shuffled, the GC content thereof, the intended degree of shuffling, etc.

Examples of the method to achieve fragmentation include use of DNase I, but other methods such as shearing by sonication can also be used. In general, fragments in the size range larger than 20 bases may be isolated from 2-3% low-melting-point agarose gel. Thus, preferably, the fragmentation may be controlled so that the average length

of the fragmented polynucleotides is from about 20 to about 500 bases. Our data showed that the size range of about 40-200 bases is further preferable for the mixing of two polynucleotide sequences of lower homology (e.g., about 80% homology), but the present invention is not limited to this range.

If regions with low sequence homologies exist, uninterrupted polynucleotide segments covering these heterogenous regions should be prepared as the hybridization and hence the recombination of two parental polynucleotide molecules would not occur at a high frequency in these regions.

Fragments of single-stranded DNA thus prepared are then used for the reassembly of fragments and amplification of assembled fragments, the steps commonly used by conventional DNA shuffling methods. Examples of family shuffling using single-stranded DNA were described below (see Examples). The frequency of the chimerical gene formation using single-stranded DNA was much higher than that using the original DNA shuffling methods with double-stranded DNAs.

In the single-stranded DNA-based methods using two family genes, single-stranded fragments of one gene could anneal only to single-stranded fragments of another gene at the 1st hybridization cycle. In the case that a mixture of

single-stranded and double-stranded polynucleotide molecules is used for the 1<sup>st</sup> hybridization cycle, (if single-stranded polynucleotide molecules are prepared by asymmetric PCR, the product is the mixture of the double-stranded and single-stranded polynucleotide molecules), the probability of the heteroduplex formation is not 100%, but it is higher than that using solely double-stranded polynucleotide molecules. Although the elongation products could make homoduplex at the 2nd or later cycles, the frequency of the heteroduplex formation would be higher in the shuffling using single-stranded DNA than that in the shuffling using double-stranded DNA.

when single-stranded DNA was used as a starting materials, the frequency of hybrid formation was 17% when two genes (*xyle* and *nahH*) having sequence divergence of about 20%. This frequency was lower than that of the restriction enzyme-based DNA shuffling in which almost all shuffling products were chimerical genes(16). However, although the restriction enzyme-based DNA shuffling has many advantages, the variety of shuffling products might be limited to a certain degree, because the variation of gene fragments generated by restriction enzyme digestion were limited compared to that generated by random fragmentation.

In the *in vitro* protein evolution, both the efficient recombination of genes and the effective screening of

recombinants are required. If a very effective screening method, such as antibiotic resistance, is available, the desired chimeras would be obtained even if the recombination efficiency is low. We could cite the selection of improved  $\beta$ -lactamase as an example of easy screens. In this case, one variant among  $10^6$  progenies could easily be selected. However, for the improvement of most industrially relevant enzymes, no selection is possible. The enzyme properties in progenies should be determined, one to the other, by using time-consuming assays. Under such circumstances, shuffling techniques with efficient recombination are of use even though these techniques require additional manipulations.

Other detailed conditions, reagents to be used, procedures, etc. for the above-described steps such as preparation of single-stranded DNA, random or non-random fragmentation of DNA molecules, annealing (hybridization), separation of double-stranded DNA into single-stranded DNA, polynucleotide synthesis (i.e., polynucleotide chain elongation) on the hybridized molecules, etc. can be appropriately decided by taking into consideration those known in the art, for example, in "Molecular Cloning, A Laboratory Manual" (edit. by T. Maniatis et al., Cold Spring Harbor Laboratory, 1989), U.S. Patent No. 5,830,721, etc., both herein incorporated by reference.

In conclusion, the present invention provides a shuffling method using single-stranded DNA which can increase the probability of the hybrid formation comparing with that using double-stranded DNA. This technique should be especially useful for the family shuffling of any genes.

#### EXAMPLE

The following results are one example of many possible embodiments of the present invention.

- (1) Family shuffling of *xyle* and *nahH* using double-stranded DNA

The *xyle* and *nahH* genes both encode catechol 2,3-dioxygenase (C230). Their nucleotide sequences are approximately 80% identical. When we applied the previously described family shuffling techniques to obtain hybrid genes of *xyle* and *nahH*, the formation of chimeric genes was severely restricted, and only the parental genes, *nahH* and *xyle*, were generated as described below.

According to the method outlined by Zhao et al. (18), we carried out the family shuffling between *xyle* and *nahH*, and the products of the family shuffling were cloned in pBluescript SK(+). After the transformation of *E. coli* cells using the ligated DNA, about 50% of the colonies developed on the selective plated showed the C230 activity. When 50 clones showing the C230 activity were randomly selected and their C230 genes were sequenced, it was found

that none of them were chimeric. Most of the products had the structure of *xylE* (26 clones), *nahH* (15 clones), or their point mutants (9 clones). Fifty additional clones were checked, however, no single chimerical gene was found. All those results demonstrated that the frequency of hybrid formation was less than 1%.

Thus, preventing the formation of the original gene structures has become quite important for successful family shuffling between *nahH* and *xylE*.

## (2) Family shuffling using single-stranded DNA

Family shuffling using single-stranded DNA were performed as described in sections (3) and illustrated in Fig. 4. After the transformation of *E. coli* cells, about 64% of the colonies grown on plates showed the C230 activity. When randomly selected 50 clones exhibiting the C230 activity were analyzed for the nucleotide sequences of their C230 genes, 7 of them (14%) were chimerical, and the other were either parental genes (40 clones) or their point mutants (3 clones). The structures of the chimerical clones were shown in Fig. 5. Two of them, hybrids 1 and 4, had the same nucleotide sequences.

The frequency of the chimerical gene formation using single-stranded DNA was much higher than that using the original sexual PCR method with double-stranded DNAs. Most likely, the probability of the homo-duplex formation in the

annealing step of 1st PCR cycle using double-stranded gene fragments may be much higher than that of the hetero-duplex formation because of the *nahH* and *xyle* sequence divergence of 20%. On the other hand, when both the single-stranded *nahH* and the single-stranded *xyle* were digested by DNase I, *nahH* gene fragments could anneal only to fragmented *xyle* at the 1st PCR cycle. Although the elongation products could make homo-duplex at the 2nd or later cycles, the frequency of the hetero-duplex formation was expected to increase in the shuffling using single-stranded DNA.

The only difference between the methods using single-stranded DNA and double stranded DNA is the template DNA for fragmentation using DNase I or other appropriate methods. For single-stranded DNA preparation, the genes have to be cloned either in a phagemid vector or in a single-stranded DNA phage such as M13, and the single-stranded DNA have to be prepared from the filamentous phage particles. Although the shuffling using single-stranded DNAs needs one more step, it is worthwhile since *nahH-xyle* chimerical genes were formed much more efficiently than in the shuffling using double stranded DNA.

(2) Detailed description of one of many possible protocols of the invention

(2-1) Preparation of ssDNAs



1. Clone two homologous genes separately in phagemid vectors (for example, pBluescript). The orientations of one gene with respect to the origin of the single-stranded DNA replication should be the same, while that of another gene should be opposite with respect to the origin of the single-stranded DNA replication. Then, the coding strand of one gene and the non-coding strand of another gene are synthesized and packaged in phage particles. In other words, single-stranded DNAs of the two homologous genes are complementary to each other.

2. Introduce the plasmids thus constructed into "male" *E. coli* such as JM109 (as the infection by single-stranded DNA phage requires F pili).

3. Inoculate a 1 ml culture of 2 X YT medium supplemented with an appropriate antibiotics with a single JM109 transformant colony.

4. Grow the culture at 37°C with shaking (e.g. rotary shaking at 250 rpm) to an optical density at 600 nm of 2.

5. Inoculate 25 ml of 2 X YT containing an appropriate antibiotics in 250 ml Erlenmeyer flask with 0.5 ml of the above culture.

6. Incubate for 1 h with shaking at 37°C.

7. Infect with a helper phage at an m.o.i. (multiplicity of infection) of 10-20. Helper phages such as

VCS-M13 (Stratagene) and M13K07 (Pharmacia) are used for the rescue of ssDNA from phagemids. These helper phages can be prepared by propagating on a male *E. coli* strain.

8. Incubate overnight with shaking at 37°C.
9. Next day, harvest the cells by centrifuging at 12,000 × *g* at 4°C for 15 min.
10. Transfer the supernatant (containing phage particles) to a fresh tube. Do not introduce the pellet in the tube.
11. Spin the supernatant again, transfer the supernatant to a fresh tube.
12. Add 0.25 volume of phage precipitation solution to the supernatant. Leave on ice for at least 1 h, or overnight at 4°C.
13. Centrifuge at 12,000 × *g* for 20 min at 4°C. In the presence of PEG, phage particles precipitate.
14. Remove the supernatant and resuspend the pellet in 400 ml of TE buffer, and transfer to a 1.5-ml tube.
15. Add one volume of TE-saturated phenol:chloroform:isoamyl alcohol (25:24:1) to the sample, vortex at least 1 min and centrifuge at 12,000 × *g* for 5 min.
16. Transfer the upper phase (containing phagemid DNA) to a fresh tube without disturbing the interface.

Repeat the organic solvent extraction until no visible material appears at the interface.

17. Add 0.5 volume (200  $\mu$ l) of 7.5 M ammonium acetate plus two volumes (1.2 ml) of 100% ethanol. Mix and leave at  $-20^{\circ}\text{C}$  for 30 min to precipitate the phagemid DNA.

18. Centrifuge at  $12,000 \times g$  for 5 min, remove the supernatant and carefully rinse the pellet with ice-cold 70% ethanol. If the pellet is disturbed, centrifuge again for 2 min. Drain the tube and dry the pellet under vacuum.

19. By agarose gel electrophoresis, two major bands corresponding to helper phage DNA and single-stranded DNA from phagemid are usually seen. A small amount of chromosomal DNA and RNA released by cell lysis may be present. Longer incubation during phagemid rescue increases the contamination of *E. coli* chromosomal DNA. Single-stranded DNA migrates faster than double stranded (ds) DNA of the same length. The presence of the helper phage DNA does not interfere with the following reactions.

#### (2-2) DNase I treatment

1. Dilute 2  $\mu$ g (or 3 pmol) of each DNA in 45  $\mu$ l of TE buffer, and add 5  $\mu$ l of  $10\times$  DNase I digestion buffer.

2. Incubate the solution at  $15^{\circ}\text{C}$  for 5 min, then add 0.3 U of DNase I.

3. Incubate further for 2 min, then transfer to  $90^{\circ}\text{C}$  and incubate for 10 min to terminate the reaction.

The concentration of DNase I, the time and temperature for the reaction should be optimized as the digestion speed of DNase I is influenced by many factors including the nature of DNA. After DNase I treatment and subsequent incubation at 90°C, the brown precipitation may be produced. This is removed by centrifugation and then the supernatant is used for PCR.

4. Fragments in the size range of 40-100 bases are isolated from 2-3% low-melting-point agarose gel using QIAEX II gel extraction kit.

5. Resuspend the recovered DNA in 50 µl of TE buffer.

#### (2-3) PCR reassembly and amplification

1. Add to a 1.5 ml tube, 10 µl of 10× Pfu DNA polymerase buffer, 10 µl of dNTP mix, 10 µl of the purified fragments, 2.5 U of Pfu Turbo DNA polymerase and H<sub>2</sub>O to the total volume of 100 µl.

2. Perform PCR: 40 cycles of denaturation at 94°C for 1 min, annealing at 56°C for 1 min and elongation at 72°C for 60 + 5 sec/cycle (the duration of elongation is increased by 5 sec after each cycle). This PCR amplification becomes difficult when the sizes of target genes are large, or the mean size of DNase I-cleave DNA is small. If the yield of the PCR products is low, the number of cycles in this PCR step should be increased.

3. Add to a new 1.5 ml tube, 1- 5  $\mu$ l of reassembled PCR product, 10  $\mu$ l of 10 $\times$  Pfu DNA polymerase buffer, 10  $\mu$ l of dNTP mix, 2  $\mu$ l of forward and reverse primers, 2.5 units of the Taq/Pfu DNA polymerase mixture and H<sub>2</sub>O to the total volume of 100  $\mu$ l.

4. Perform PCR: 96°C for 2 min followed by 25 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 45 + 20 sec/cycle. This PCR amplification using primers seems to be inhibited when the concentrations of reassembled PCR products (i.e. the product of the first PCR without primers: step 2) are high. Check the concentrations of the first (step 2) and second (this step) PCR products on an agarose gel. If the second PCR amplification is not efficient, reduce the concentration of substrates (i.e. the product of the first PCR) in the second PCR.

5. Separate the PCR products on an agarose gel and recover a band corresponding to the size of the full length gene.

6. Purify the PCR products, subclone them and select desired mutants among transformants.

#### (2-4) Isolation of thermally stable C230s

Thermally stable clones were screened from 750 colonies obtained by single-stranded DNA-based shuffling. After the treatment at 65°C for 10 min, under which *Xyle* (the *xyle* gene product, namely C230 produced from *xyle*) and

NahH (the *xyle* gene product, namely C230 produced from *nahH*) were inactivated, 10 colonies exhibited the residual C230 activity, showing that they contain enzymes thermally more stable than the wild-type C230s. The amino acid sequences of the thermally stable C230s are shown in Fig. 6. Although the nucleotide sequences of all 10 clones were different, the deduced amino acid sequences of clones 120 and 942, clones 202 and 450, and clones 315 and 1527 were identical.

#### (2-5) Notes concerning DNA polymerases

Primer design is crucial in DNA shuffling. Primer length and sequence should carefully be determined for a successful amplification, cloning and gene expression. Follow general guidelines on the primer design for PCR cloning that are described in many textbooks. Sequences flanking the target gene can also be used for the primer design.

Taq DNA polymerase adds a single 3'-dA overhang to a blunt dsDNA template. For the cloning of such PCR products, linearized plasmid vectors containing single dT overhangs (T-vectors) were developed. The cloning efficiencies of the T-vectors, however, were variable. DNA polymerases with 3'→5' exonuclease (proofreading) activity remove mispaired nucleotides from 3' ends of dsDNA and generate blunt-end PCR products. The PCR products

generated by these proofreading polymerases thus can be cloned into vectors by blunt-end ligation. However, blunt-end cloning of PCR products is less efficient than sticky-end cloning. Thus, it is advisable to introduce additional restriction sites at the 5' end of each of the primers. As amplification proceeds, these primers are incorporated into the PCR product. Thus, the PCR products can be digested by appropriate restriction enzymes to clone in an appropriate vector.

Taq polymerase lacks a 3'→5' exonuclease activity. In other words, it lacks the proofreading function of DNA replication, and thus exhibits a high rate of replication errors. This enzyme shows higher replication errors when the concentration of  $MnCl_2$  in the reaction mixture increased, or when the concentration of one nucleotide was lower than those of other three nucleotides in the reaction mixture. By this reason, PCR with Taq polymerase was often used for the mutagenesis of genes. Pfu and Pfu Turbo DNA polymerases, on the other hand, show high fidelity of DNA replication. Zhao *et al.* (18) have reported that the using a proofreading DNA polymerase (Pfu or Pwo) in the reassembly step increased the frequency of active clones from the shuffling products. Therefor PCR with the proofreading DNA polymerase (Pfu Turbo) is used here to reduce the negative mutation.

The amplification by Pfu Turbo polymerase is not as powerful as that by Taq DNA polymerase. The replacement of Pfu Turbo polymerase by Taq DNA polymerase or by the mixture of Pfu Turbo and Taq polymerases often provides better results.

References:

The references cited above are shown below, the entire contents of each being hereby incorporated by reference.

1. Stemmer, W.P.C. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc. Natl. Acad. Sci. U.S.A., 91, 10747-10751.
2. Stemmer, W.P.C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature (London), 370, 389-391.
3. Shao, Z., Zhao, H., Giver, L., and Arnold, F.H. (1998) Random-priming in vitro recombination: an effective tool for directed evolution. Nucleic Acids Res., 26, 681-683.
4. Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat. Biotechnol., 16, 258-261.



5. Cramer, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P.C. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.*, 14, 315-319.

6. Cramer, A., Dawes, G., Rodriguez, E., Jr., Silver, S., and Stemmer, W.P.C. (1997) Molecular evolution of an arsenate detoxification pathway by DNA shuffling. *Nat. Biotechnol.*, 15, 436-438.

7. Moore, J. C. et al. (1997) Strategies for the in vitro evolution of protein function: Enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.* 272, 336-347.

8. Yano, T., Oue, S., and Kagamiyama, H. (1998) Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U.S.A.*, 95, 5511-5515.

9. Zhang, J.H., Dawes, G., and Stemmer, W.P.C. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. U.S.A.*, 94, 4504-4509.

10. Patten, P.A., Howard, R.J., and Stemmer, W.P. (1997) Applications of DNA shuffling to pharmaceuticals and vaccines. *Curr. Opin. Biotechnol.*, 8, 724-733.

11. Cramer, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P.C. (1998) DNA shuffling of a family of genes

from diverse species accelerates directed evolution.  
Nature (London), 391, 288-291.

12. Harayama, S. (1998) Artificial evolution by DNA shuffling. Trends Biotechnol., 16, 76-82.

13. Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T., and Furukawa, K. (1998) Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nat. Biotechnol., 16, 663-666.

14. Chang, C.-C., Chen, T.T., Cox, B.W., Dawes, G.N., Stemmer, W.P.C., Punnonen, J., and Patten, P.A. (1999) Evolution of a cytokine using DNA family shuffling. Nat. Biotechnol., 17, 793-797.

15. Hansson, L. O., B-Grob, R., Massoud, T., and Mannervik, B. (1999) Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling. J. Mol. Biol., 287, 265-276.

16. Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) Novel family shuffling methods for the in vitro evolution of enzymes. Gene, 236, 159-167.

17. Kikuchi, M., Ohnishi, K., and Harayama, S. (2000) An effective family shuffling method using single-stranded DNA. Gene, 243, 133-137.

18. Zhao, H., and Arnold, F.H. (1997) Optimization of DNA shuffling for high fidelity recombination. Nucleic Acids Res., 25, 1307-1308.

While the invention has been described in detail and with reference to specific embodiments thereof, it will be apparent to one skilled in the art that various changes and modifications can be made therein without departing from the spirit and scope thereof.